

## CLAIMS

What is claimed is:

- 1 1. A method comprising:
  - 2 receiving diphone waveforms;
  - 3 compressing the diphone waveforms into diphone residuals, wherein the
  - 4 compressing is performed using an encoder;
  - 5 generating linear predictive coding (LPC) coefficients, wherein the LPC
  - 6 coefficients are generated by the encoder; and
  - 7 storing the diphone residuals and the encoder-generated LPC coefficients in a
  - 8 compressed packet, wherein the compressed packet is generated by the
  - 9 encoder.
- 1 2. The method of claim 1 further comprising:
  - 2 a waveform synthesizer requesting diphone residuals;
  - 3 locating the requested diphone residuals in the compressed packet;
  - 4 extracting the located diphone residuals from the compressed packet;
  - 5 decompressing the extracted diphone residuals, wherein the decompressing is
  - 6 performed using a decoder; and
  - 7 supplying the diphone residuals to the waveform synthesizer.
- 1 3. The method of claim 2 further comprising supplying the encoder-generated LPC
  - 2 coefficients to the waveform synthesizer.

1 4. The method of claim 2 further comprising supplying pitch marks to the waveform  
2 synthesizer;

1 5. The method of claim 2 further comprising the waveform synthesizer producing  
2 speech output.

1 6. The method of claim 1, wherein the encoder is a G.723 encoder.

1 7. The method of claim 1, wherein the decoder is a modified G.723 decoder.

1 8. A method comprising:

2 receiving diphone waveforms;

3 compressing the diphone waveforms into diphone residuals, wherein the

4 compressing is performed using an encoder;

5 generating linear predictive coding (LPC) coefficients, wherein the LPC

6 coefficients are generated by the encoder;

7 storing the diphone residuals and the coder-generated LPC coefficients in a

8 compressed packet, wherein the compressed packet is generated by the

9 encoder;

10 a waveform synthesizer requesting the diphone residuals;

11 locating the requested diphone residuals in the compressed packet;

12 extracting the located diphone residuals from the compressed packet; and

13 decompressing the extracted diphone residuals, wherein the decompressing is

14 performed using a decoder; and

15 supplying the diphone residuals and the encoder-generated LPC coefficients to the  
16 waveform synthesizer.

1 9. The method of claim 8 further comprising supplying pitch marks to the waveform  
2 synthesizer.

1 10. The method of claim 8, wherein the encoder is a G.723 encoder.

1 11. The method of claim 8, wherein the decoder is a G.723 decoder.

1 12. A system for compressing and using concatenative speech databases in text-to-  
2 speech systems comprising:

3 a text-to-speech system;

4 a concatenative speech database; and

5 a coder.

1 13. The system of claim 12, wherein the text-to-speech system comprising:

2 a text analysis module for processing a text into forms of linguistic

3 representations;

4 a linguistic and prosodic analysis module for analyzing the forms of linguistic

5 representations corresponding to their assigned language system; and

6 a waveform synthesizer for producing a speech output.

1 14. The system of claim 12, wherein the concatenative speech database comprising:

2 diphone waveforms;

3 LPC coefficients; and

4 pitch marks.

- 1 15. The system of claim 14, wherein the diphone waveforms are compressed to  
2 diphone residuals.
- 1 16. The system of claim 12, wherein the coder is a G.723 coder.
- 1 17. The system of claim 16, wherein the G.723 coder comprises:  
2 a G.723 encoder for compressing the concatenative speech database; and  
3 a G.723 decoder for decompressing the concatenative speech database.
- 1 18. A method of producing a compressed concatenative diphone database comprising:  
2 compressing diphone waveforms and generating linear predictive coding (LPC)  
3 coefficients by applying an audio encoder to the diphone waveforms; and  
4 storing compressed packets produced by the audio encoder and uncompressed  
5 pitch mark values as a compressed concatenative diphone database.
- 1 19. The method of claim 18, wherein the compressed packets comprising diphone  
2 residuals and audio encoder-generated LPC coefficients.
- 1 20. The method for a handheld device with a text-to-speech system using a  
2 compressed concatenative diphone database comprising:  
3 compressing diphone waveforms into diphone residuals and generating linear  
4 predictive coding (LPC) coefficients by applying an audio encoder to the  
5 diphone waveforms;  
6 storing compressed packets produced by the audio encoder and uncompressed  
7 pitch mark values as a compressed concatenative diphone database;

8 decompressing the compressed concatenative diphone database by applying an  
9 audio decoder to the diphone residuals and the LPC coefficients; and  
10 synthesizing the decompressed concatenative diphone database including the  
11 uncompressed pitch mark values to produce an output by applying a  
12 waveform synthesizer.

1 21. The method of claim 20 further comprising the handheld device downloading a  
2 customizable speech database.

1 22. The method of claim 20, wherein the synthesizing is client-based.

1 23. A concatenative speech database structure comprising:  
2 diphone waveforms indicating smallest units of speech for efficient text-to-speech  
3 conversion that are derived from phonemes;  
4 linear predictive coefficients of a difference equation for characterizing formants;  
5 and  
6 pitch mark values marking positions in an utterance indicating varying pitch.

1 24. The concatenative speech database structure of claim 23, wherein the diphone  
2 waveforms are reduced to diphone residuals after compression.

1 25. The concatenative speech database structure of claim 23, wherein the difference  
2 equation is a linear predictor expressing each new sample of a signal as a linear  
3 combination of previous samples.

1 26. The concatenative speech database structure of claim 23, wherein the formants are  
2 the resonance characterizing vocal tract.

